**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**

**FACULTAD DE CIENCIAS SOCIALES**



**ENSAYO**

Tipula Palli, Diego (20162580)

**Curso**

Estadística para el análisis político 2

**Docente**

José Manuel Magallanes Reyes

**Jefe de práctica**

Wendy Adrianzén Rossi

Lima, 2024

**Introducción**

Actualmente, es notable que persisten los casos de corrupción a nivel global. Estos pueden ser percibidos incluso desde la experiencia personal o en la vida cotidiana. Así, hechos como la corrupción política son muy comunes en cualquier país o Estado. Ello es claro, ya que “Por doce años consecutivos, el promedio global del IPC se mantiene sin variaciones en 43 puntos, y más de dos tercios de los países obtienen una puntuación inferior a 50” (Transparency International, 2023). Esto indica que existen graves problemas de corrupción”.

En tal sentido, el tema del siguiente ensayo tratará de analizar sobre los factores que influyen en el índice de corrupción política. Cabe añadir que se escogió este tema porque representa gran relevancia a nivel teórico y práctico; en efecto, **abordar dichas variables resulta útil a nivel práctico, ya que, una vez comprendida su relación e influencia, se pueden generar medidas, estrategias o políticas para reducir a gran escala la corrupción en cada país. Además,** a nivel teórico posee relevancia mediante el uso de las variables como el índice de corrupción, GDP per capita, tasa de desempleo, tasa de libertad de prensa y si posee o no leyes transparentes. Así, contribuyen a explicar de qué manera y por qué la corrupción sucede. **En vista de ello,** Vera (2015) plantea que “[…] el tratamiento de estos problemas requiere tratar las causas económicas, políticas e institucionales subyacentes y, sobre todo, culturales. Es decir, existen acciones que, en conjunción con las ya existentes (leyes de acceso a la información y mecanismos de rendición de cuentas) pueden ser viables para un impacto positivo en la evaluación de la gobernanza y disminución de la corrupción del Estado”. Por ello, el índice de democracia, la tasa de desempleo, el índice de libertad de prensa y el nivel de leyes transparentes pueden funcionar como variables predictoras, ya que abarcan el espacio institucional y político. Asimismo, según Rose-Ackerman (2001), “dado que los índices de corrupción se corresponden mucho con otras medidas de eficiencia burocrática, tales como el nivel de burocratizados y la calidad de los jueces, estos estudios no logran distinguir el efecto marginal que puede tener cualquiera de estas medidas si mantenemos las demás constantes”. En ese sentido, la autora propone que tales variables que incluyen la burocracia o temas relacionados por sí solos podría presentar dificultades al momento de explicar la relación con la variable dependiente propuesta. En ese sentido, variables como el GDP per cápita puede contribuir al análisis estadístico. Así, si se quisiera analizar la relación entre las independientes y el índice de corrupción política, la inclusión de esta variable de control puede ayudar a verificar si la relación persiste o no.

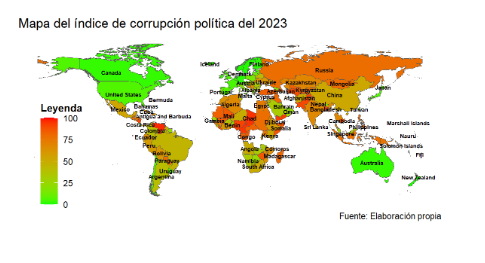
Por todo ello, queda claro que mi pregunta de investigación es **sobre qué factores explican las variaciones en el índice de corrupción política observadas en los países y Estados no reconocidos a nivel mundial en 2023. De tal manera, mi hipótesis plantea que, a priori, todas las variables tienen un impacto significativo sobre el índice de corrupción, de modo que disminuyen el índice de corrupción política.**

**Metodología y data**

Respecto a las variables, tanto la variable dependiente como las independientes son numéricas y continuas a excepción de la variable “Transp Laws” la cual es dicotómica. Cada una de ellas será usada tal cual y no será necesaria ninguna recategorización. No obstante, solo en el caso de la variable dependiente “Pol\_corruption\_index” su rango que varía entre 0 a 1 será reemplazado proporcionalmente al rango que varía entre 0 a 100 con fines de que los gráficos sean atractivamente más visuales. Además, como ya se mencionó, se hará uso de una variable de control la cual representa el GDP per cápita cuyo valor es numérico y continuo. En cuanto a la obtención de los índices y tasas de las variables mencionadas, cada una de ellas proviene de la página “Our World in Data”, ya que es lugar rico en datos. Asimismo, facilitó el hecho de que cada data encontrada está en inglés por lo que no hubo problemas con usar la función “merge”. Sobre la limpieza, la variable dependiente es usada sobre el actual índice de corrupción; es decir, se filtró el índice de 2023. Así, se tomó en cuenta de que ninguna sobrepasara dicho año para no generar algún sesgo. En efecto, las independientes son de 2023, 2022 o 2021; de tal manera, cada una pertenece a cualquiera de estos 3 años en su totalidad de la variable por lo que proporciona un análisis más actual.

**Hallazgos**

En cuanto al análisis univariado, se hizo uso gráficos y tablas como el mapa de Indice de corrupción política o la tabla con medidas de tendencia central y dispersión respecto a dicha variable dependiente.

**Gráfico 1.**

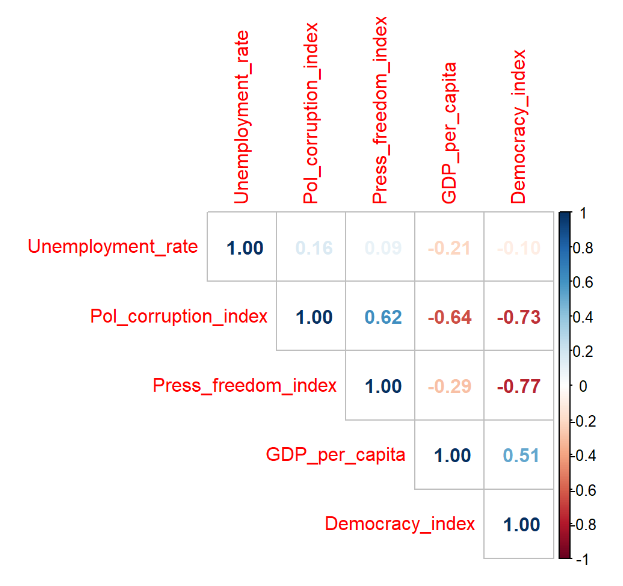
Se puede observar que mayormente existen tonalidades rojizas, lo cual indica que puede haber más casos de países con un mayor índice de corrupción a comparación de los que presentan poco.

**Cuadro 1.**



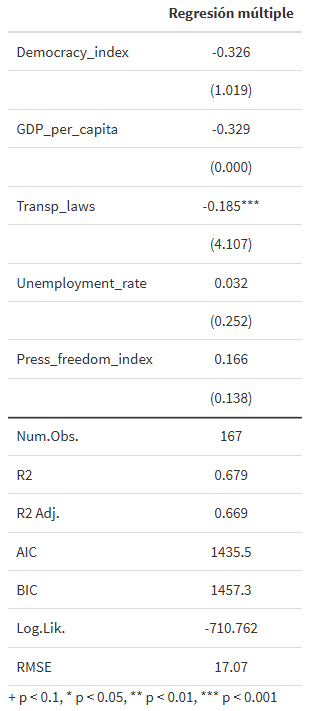
Fuente: elaboración propia

A partir del cuadro con tendencias centrales y dispersión, se puede observar que el caso mínimo es 0.3 y el caso máximo representa 96.7. Asimismo, se puede evidenciar que tanto la media como la mediana son prácticamente iguales. Es decir, la mayoría de datos parecen estar centradas en dicho valor. Cabe añadir que no existe una asimetría preocupante. No obstante, parece haber cierta variabilidad en cuanto al IQR; es decir, hay dispersión entre el Q1 y Q3.

**Gráfico 2.**

Fuente: elaboración propia

En el gráfico 2 se puede evidenciar un “Corrplot” con las correlaciones entre todas las variables; no obstante, es preciso centrarnos en las correlaciones de la variable dependiente con cada una de las demás. En efecto, se obtiene valores como 0.62, -0.64, -0.73 y 0.16. Respecto a este último valor de 0.16, representa la correlación entre la tasa de desempleo y el índice de corrupción el cual es la correlación más débil en relación a las demás ligadas a la variable dependiente. Según el criterio de Cohen, todas las demás resultan fuertes independientemente de si son negativas o positivas.

**Cuadro 2.**

Fuente: elaboración propia

Como ya se había dicho anteriormente, la intención era hacer una regresión múltiple para evaluar el impacto en conjunto sobre la variable dependiente. Además, se optó por estandarizar los coeficientes de manera apropiada al ser que algunas variables tenían escalas distintas. Ello se logró por medio del paquete “lm.beta”. Además, al ser un modelo válido y siguiendo nuestro intervalo, diremos que el modelo tiene un ALTO nivel explicativo, dado que el R2 ajustado es de 0.669. Así, 66.90% de la variabilidad de Y es explicada por nuestras X´s. Vemos que la variable que tendría mayor impacto es “Transp Laws”. En tal sentido, sí tiene efecto significativo al 0.001, aunque no es directo, puesto que el coeficiente calculado es negativo. En cuanto a su magnitud de ese efecto es -0.185, lo que indica cuanto varía el índice de corrupción en promedio cuando “Transp\_Laws” es 1 y no 0 controlado la variable de control. Se puede decir que, entre todas las variables analizadas, “Transp\_laws” es la que más contribuye a explicar la variabilidad en el índice de corrupción. Esto sugiere que las leyes de transparencia tienen un impacto importante y significativo en la reducción de la corrupción. Por otra parte, el modelo cumple con todos los diagnósticos de regresión. Es decir, cumple con el supuesto de linealidad, homocedasticidad, normalidad de residuos, no multicolinealidad y valores influyentes**[[1]](#footnote-1)**.

Es decir, sigue una relación lineal. Del mismo modo, es homocedástico y los residuos se distribuyen de manera normal lo cual es bueno para la regresión. Cabe resaltar que no existe multicolinealidad, de modo que no existe una gran correlación entre todas las variables, de modo que no se pueden agrupar en un índice las independientes. Además, no presenta valores influyentes, por lo que no se afectan los cálculos de la regresión.

**Gráfico 3.**

Fuente: elaboración propia

Por otro lado, se optó por realizar varios métodos de clusterización. De tal modo, según el gráfico 5, el método por DIANA resulta la más apropiada en términos de precisión. Así, posee una estructura no tan compleja, pero con segmentaciones necesarias para poder comprender las diferencias entre las observaciones. Del mismo modo, se puede observar que la mayoría de observaciones se agrupan en el cluster 3 y existe solamente un valor mal clusterizado a comparación del método AGNES y PAM los cuales presentan muchos más.

**Conclusiones**

A partir del análisis de la variable central, se concluye que, a pesar de que la media y la mediana sean iguales, existe variabilidad en cuanto al IQR, de modo que hay dispersión entre el primer y tercer cuartil.

Por otro lado, luego de analizar la variable dependiente en función de las independientes por medio de una regresión lineal múltiple, se puede concluir que el modelo de regresión es válido. Además, posee un alto nivel explicativo, ya que su R2 ajustado determina que 73.81% de la variabilidad de la dependiente es explicada por las independientes.

Del mismo modo, se concluye que la variable predictora que tiene más efecto sobre la dependiente del índice de corrupción es “Transp Laws”. En otras palabras, el índice de corrupción varía en -0.185 en promedio cuando dicha variable es 1 y no 0.

No obstante, puesto que las otras variables predictoras no resultaron significativas, es posible que existan otras variables que no se consideraron y que podrían tener cierto efecto. Sin embargo, a partir de este análisis, queda claro que dichas variables predictoras no tienen impacto sobre el índice de corrupción y no deberían ser tomadas en cuenta en un análisis similar, a pesar del hecho de haber cumplido con los supuestos de regresión.

Por último, se concluye el hecho de que se ha determinado la clusterización por medio de 3 clusters, lo cual sugiere que existen varias categorías distintas de países con diferentes perfiles de corrupción y características asociadas.

**BIBLIOGRAFÍA**

*Índice de Percepción de la Corrupción 2023: El debilitamiento de los sistemas de justicia deja a la corrupción sin controles - Press*. (2024, enero 30). Transparency.org. <https://www.transparency.org/es/press/cpi2023-corruption-perceptions-index-weakening-justice-systems-leave-corruption-unchecked>

Rose-Ackermann, S. (2001). Desarrollo y corrupción. *Gestión Y Análisis De Políticas Públicas*, (21), 5–21. <https://doi.org/10.24965/gapp.vi21.260>

Vera Martínez, M. C. (2015). El acceso a la información pública y su impacto en la corrupción. El caso de México. *Gestión Y Análisis De Políticas Públicas*, (13), 45–56.

<https://doi.org/10.24965/gapp.v0i13.10230>

1. DashboardCorruption

   <https://github.com/4ntonnn/EAP2/blob/main/DashboardCorruption.html> [↑](#footnote-ref-1)